1   **Supplementary information to the following publication:** Griffith, D.M., Byrd, K.B.,
2   Anderegg, L.D.L., Yacoub, R., Gatziolis, D., Roberts, D., Nemani, R.R. (2022) Capturing
3   patterns of evolutionary relatedness with reflectance spectra to model and monitor biodiversity.
4
5   **Contents**
6
7   Supplementary Figure 1
8   Supplementary Figure 2
9   Supplementary Figure 3
10 Supplementary Figure 4
11 Supplementary Figure 5
12 Supplementary Figure 6
13 Supplementary Methods 1
14 Supplementary Figure 8
15 Supplementary Figure 9
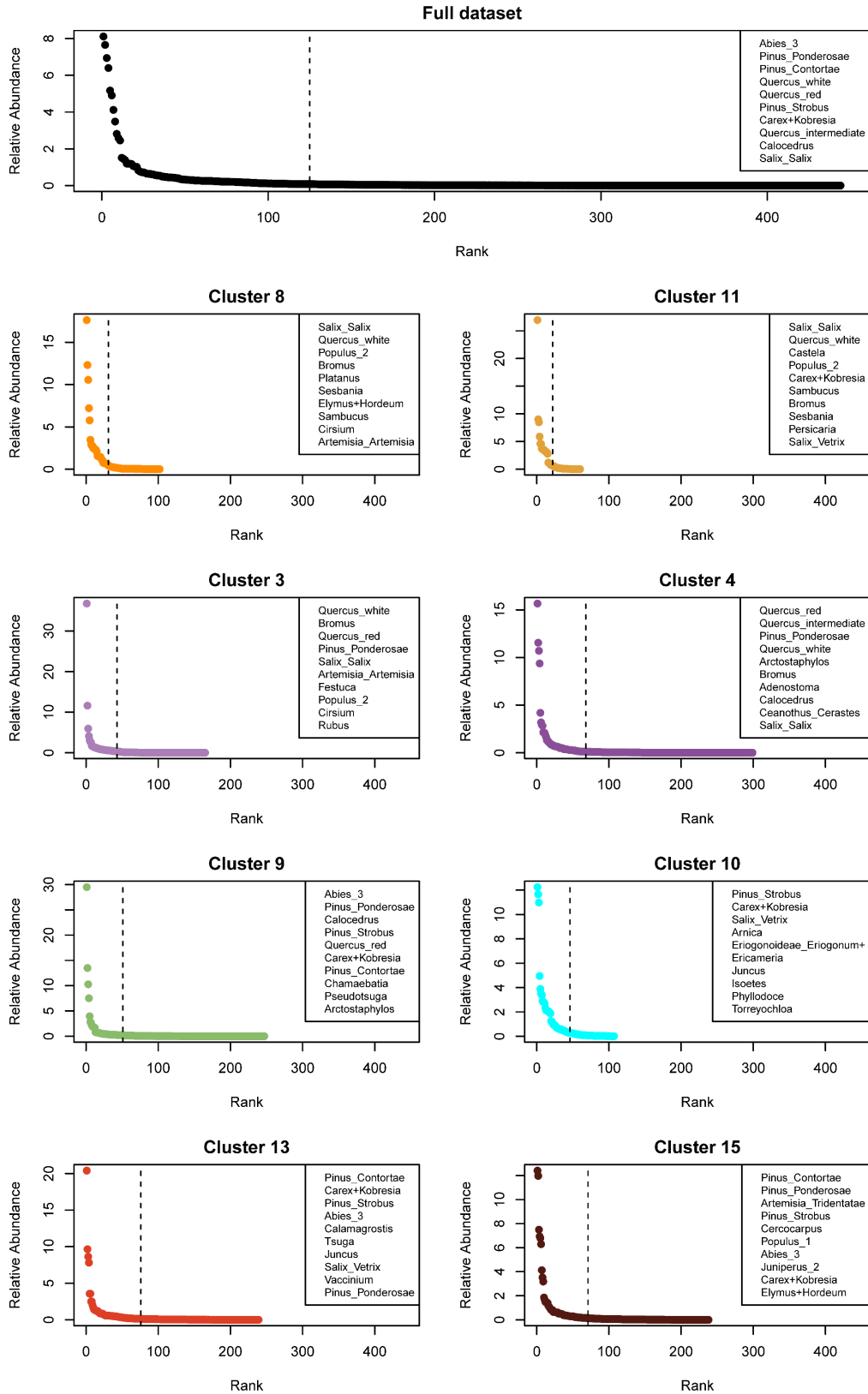16 Supplementary Figure 10
17
18
19
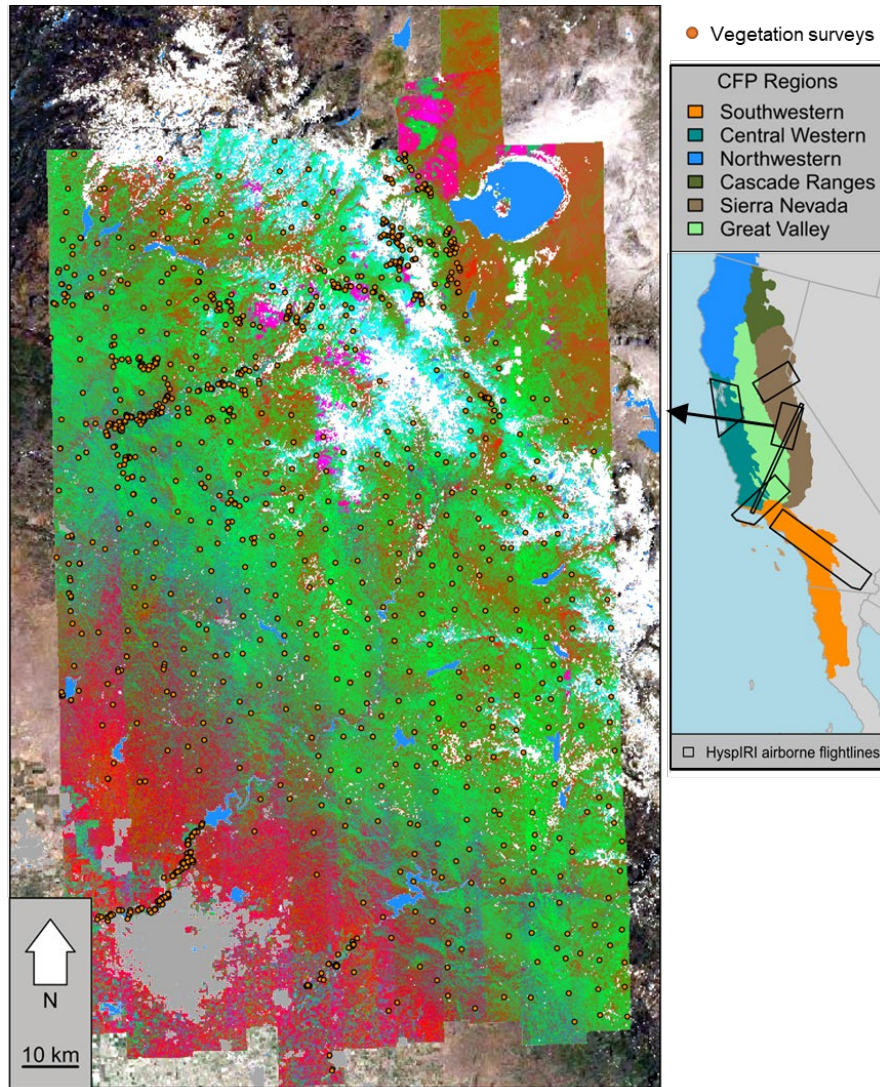
20
21
22
23   Figure S1. Minimum Noise Fraction representation (bands 2, 3, and 4 as RGB) of the Southern
24   Sierra Nevada Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) hyperspectral data.
25   This visualization approach illustrates the variation present in the highly multi-dimensional
26   spectral dataset. Forest Inventory and Analysis (FIA) plot data are included in this figure
27   (locations for figure are spatially fuzzed following FIA protocol, but exact plot locations were
28   used in analysis). Notably, 'banding effects' due to influences such as sun-angle effects are
29   highly noticeable in the spectral data along each flight line. We discuss the impact of these
30   artifacts for our study and remote sensing in the Discussion.
31
32
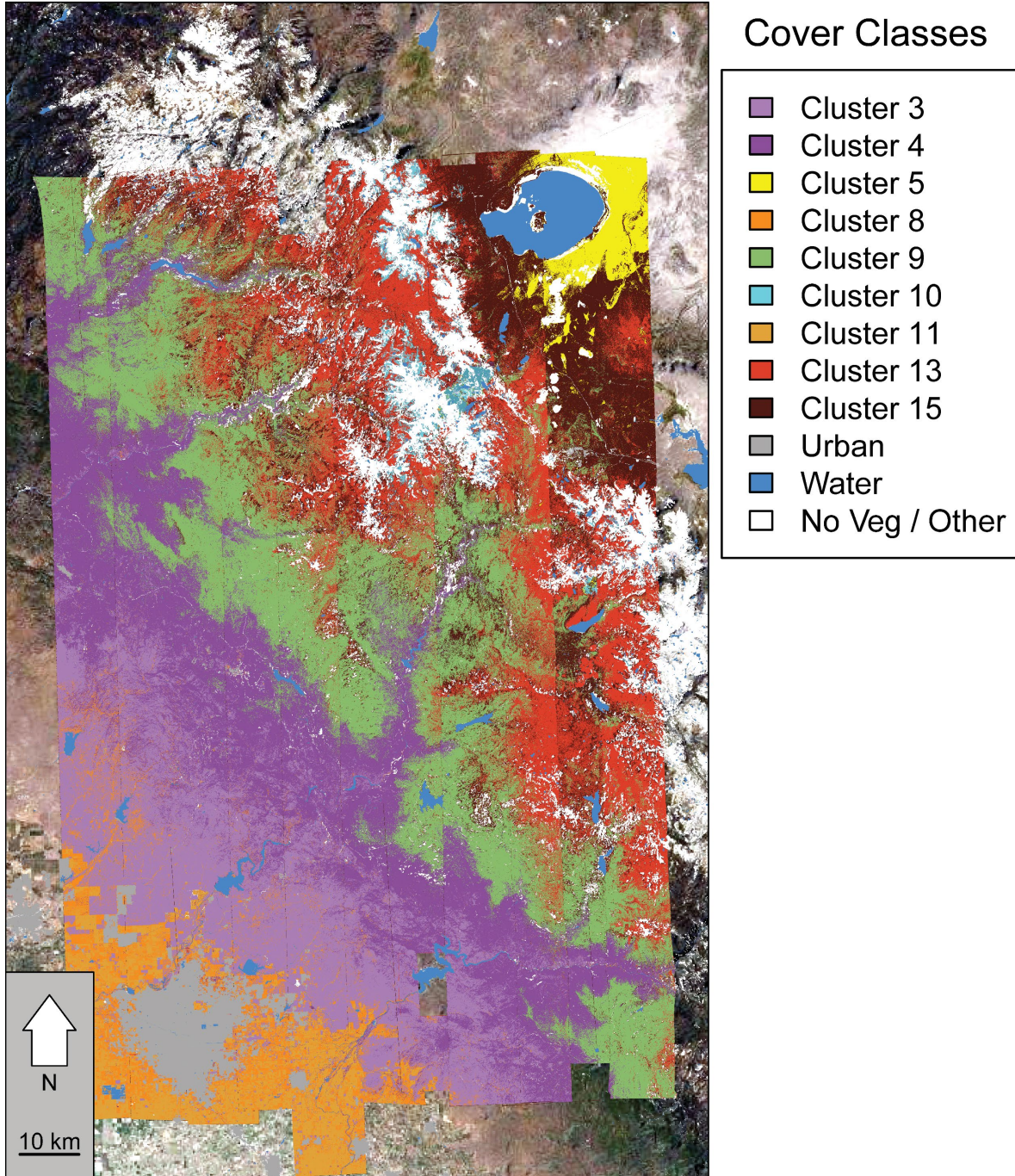
33
34
35 Figure S2. Results of K-means clustering of spectra. Spectral Clusters 8 and 11 are associated
36 with agricultural activity around Fresno, CA, and intersected 69 and 28 vegetations plots
37 respectively (they were assigned two shades of orange). Spectral Clusters 3 and 4 had 99 and 274
38 plots. Spectral Clusters 9, 10, 13, 15 had 222, 35, 260, and 165 plots, respectively. Sun-angle
39 impacts are noticeable and discussed in text.
40
41

Full dataset

Abies_3
Pinus_Ponderosae
Pinus_Contortae
Quercus_white
Quercus_red
Pinus_Strobus
Carex+Kobresia
Quercus_intermediate
Calocedrus
Salix_Salix

Cluster 8

Salix_Salix
Quercus_white
Populus_2
Bromus
Platanus
Sesbania
Elymus+Hordeum
Sambucus
Cirsium
Artemisia_Artemisia

Cluster 11

Salix_Salix
Quercus_white
Castela
Populus_2
Carex+Kobresia
Sambucus
Bromus
Sesbania
Persicaria
Salix_Vetrix

Cluster 3

Quercus_white
Bromus
Quercus_red
Pinus_Ponderosae
Salix_Salix
Artemisia_Artemisia
Festuca
Populus_2
Cirsium
Rubus

Cluster 4

Quercus_red
Quercus_intermediate
Pinus_Ponderosae
Quercus_white
Arctostaphylos
Bromus
Adenostoma
Calocedrus
Ceanothus_Cerastes
Salix_Salix

Cluster 9

Abies_3
Pinus_Ponderosae
Calocedrus
Pinus_Strobus
Quercus_red
Carex+Kobresia
Pinus_Contortae
Chamaebatia
Pseudotsuga
Arctostaphylos

Cluster 10

Pinus_Strobus
Carex+Kobresia
Salix_Vetrix
Arnica
Eriogonoideae_Eriogonum+
Ericameria
Juncus
Isoetes
Phyllodoce
Torreyochloa

Cluster 13

Pinus_Contortae
Carex+Kobresia
Pinus_Strobus
Abies_3
Calamagrostis
Tsuga
Juncus
Salix_Vetrix
Vaccinium
Pinus_Ponderosae

Cluster 15

Pinus_Contortae
Pinus_Ponderosae
Artemisia_Tridentatae
Pinus_Strobus
Cercocarpus
Populus_1
Abies_3
Juniperus_2
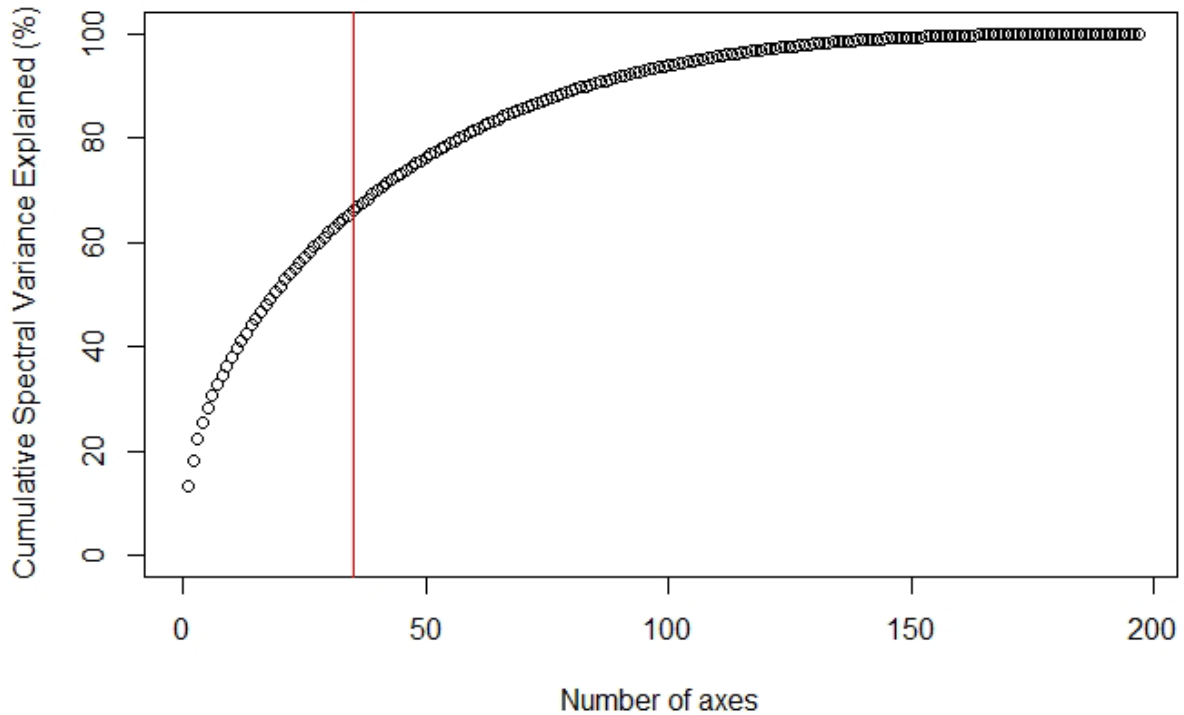Carex+Kobresia
Elymus+Hordeum

42
43   Figure S3. Rank abundance curves for vegetation plot data in each of 8 spectral clusters. The
44   vegetation data come from Vegbank, VegCamp, and FIA.
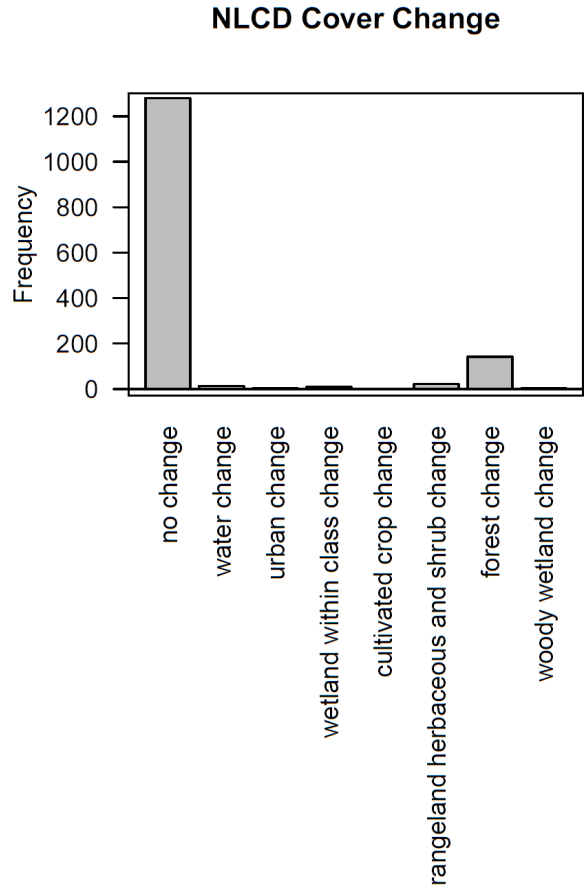
NLCD Cover Change

45
46   Figure S4. Frequency of original dataset in each land cover change category from years 2000 to
47   2019.
48
49
50

51
52 Figure S5. Selection of CDA (Canonical Discriminant Analysis) axes for dimension reduction.
53 This screeplot represents the cumulative variance explained in the spectral data based on the
54 number of CDA axes. The axes in CDA are optimized for emphasizing spectral difference
55 among the dominant plant species in each represented spectrum based on vegetation data, which
56 differs from more standard ordination approaches. As a result, the screeplot is not as steep as
57 many other methods of ordination as the focus is not on explaining spectral variation with few
58 axes, but instead on distinguishing plants. We selected the axes that together explained two-
59 thirds of the variation in spectral data based on the number of bands selected in Meerdink et al
60 (2019) and to enable data processing to occur in a reasonable time frame.

61
62
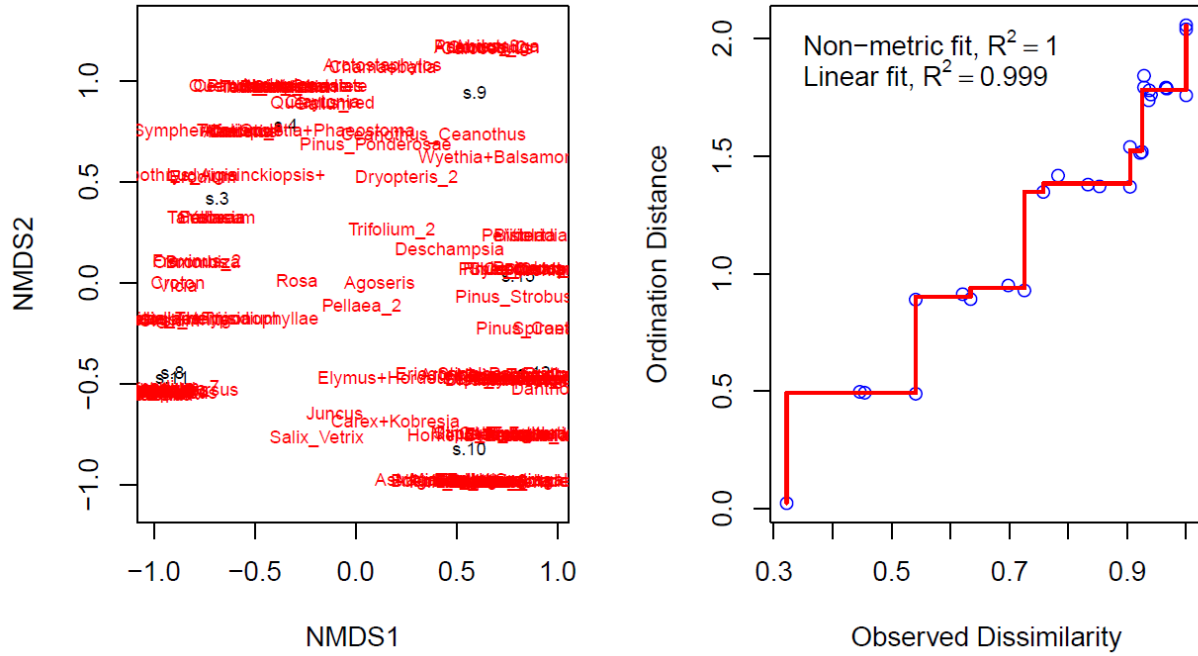
63
64
65
66 Figure S6. Non-metric Multidimensional Scaling ordination results. On the left panel, red text
67 shows the Operational Taxonomic Units (OTU) positions in in multivariate space, whereas black
68 text indicates the position of Spectral Clusters in multivariate space (e.g., "s.9" is Spectral
69 Cluster #9).
70
71
72
73
74
75
76
77
78
79

80 Supplementary Methods S1.
81 We considered several approaches for generating Lineage Functional Types. First, we considered
82 a simple age-based approach that cut the phylogeny into different sized lineages using an age
83 threshold (Fig. S7). However, this approach was unsatisfactory because it was biased towards the
84 basal grade and had no potential for explicitly including abundance data or for vegetation traits
85 that should be balanced in the generation of LFTs. Alternatively, we tried the KSI method from
86 Cornwell et al., (2014) which searches for lineages with distinct functional synapomorphies.
87 While a modification of this approach might be a major improvement in the future, the result we
88 received with PCA axes as substitutes for traits didn't reflect the vegetation patterns we see in
89 the Indicator Species Analysis and it selected lineages that were most unique (leaving many
90 species unselected) and not balanced across the tree. Fig S8 illustrates the poor performance of
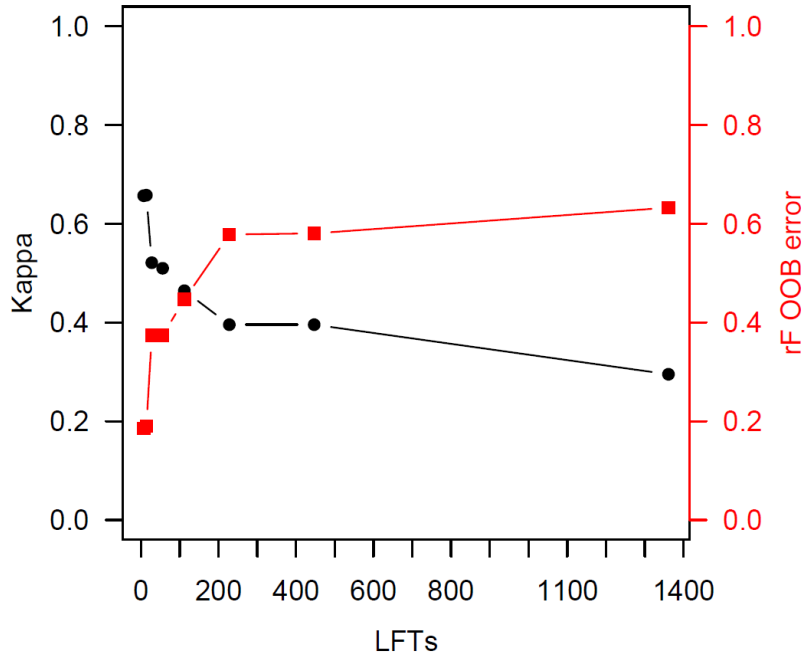91 this method as compared to the simple spectrally informed LFT generation approach we used.
92
93
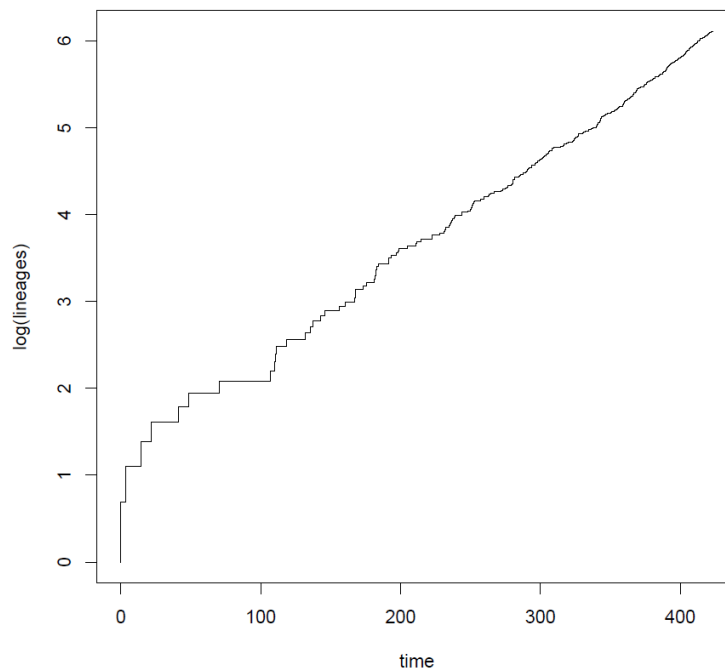
94
95
96
97 Figure S7. Log lineages over time, used for splitting the tree based on age. These cuts were
98 completed on the full tree which comprises the 1572 species in our study.
99
100

101
102
103   Figure S8. PLS-DA classification Kappa in black and random forest Out of Bag (OOB) error in
104   red for classifications based on different numbers of cuts to phylogeny.
105
106
107
108
109
110
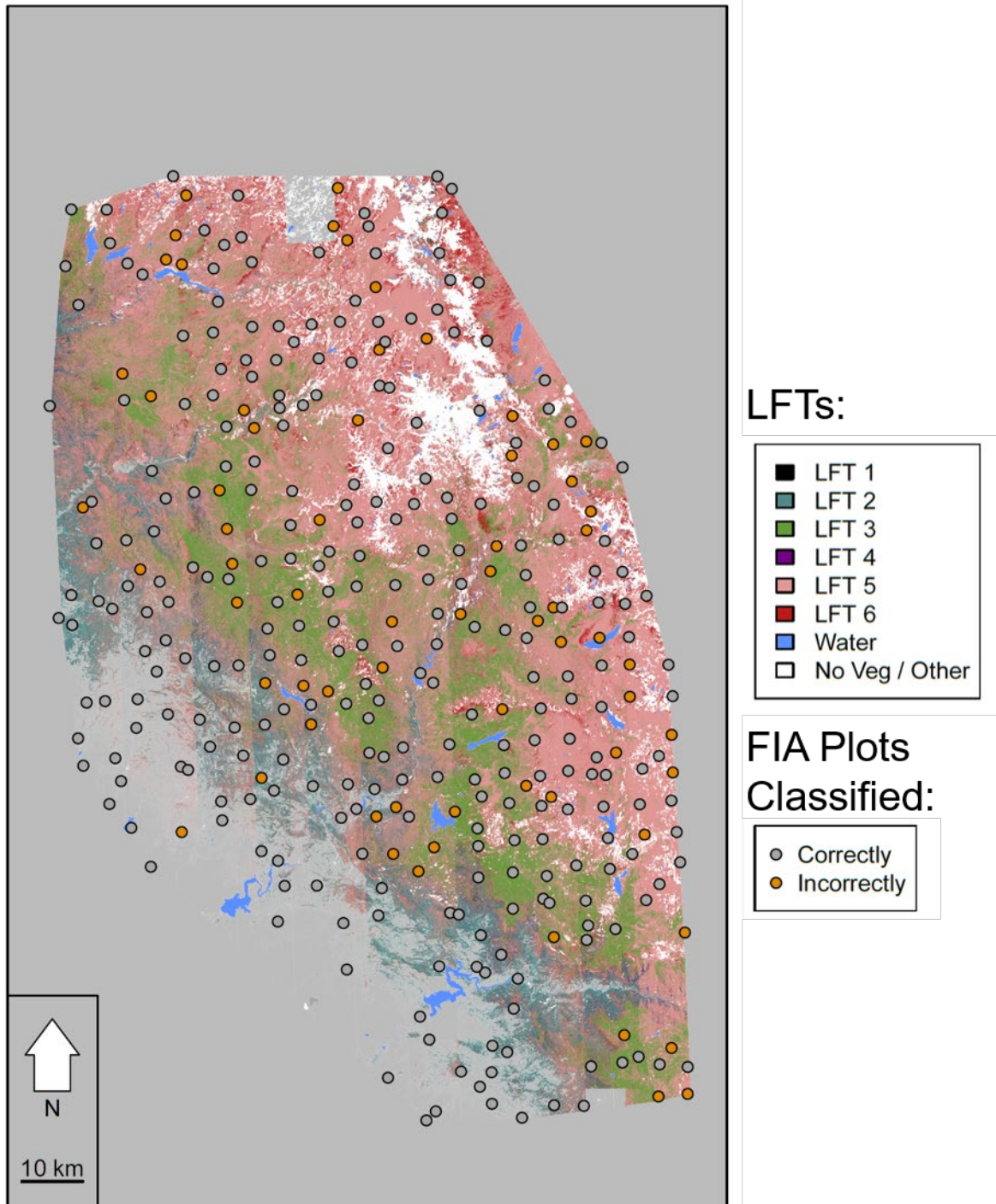
111
112   Figure S9. Spatial visualization of Support Vector Machine classification error based on the FIA
113   data. FIA plots shown in the figure are fuzzed-and-swapped but the spatial statistics were
114   performed using the actual plot coordinates. We used the Join-Count Statistic in the spdep R
115   package to show that the classification errors are not clustered (p = 0.43).
116

117
118
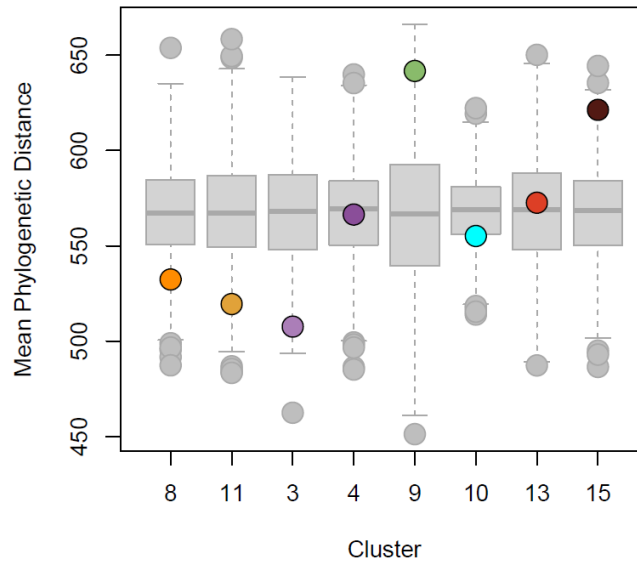
119
120   Figure S10. Mean Phylogenetic Distance (MPD) for each spectral cluster overlayed on null
121   distributions generated by simulating community assembly from the phylogeny in Fig. 2 999
122   times. Over half of the naïve spectral clusters represented indicator species that were more
123   phylogenetically related than expected, and only two of the eight clusters were phylogenetically
124   'over dispersed,' or less related than expected. We calculated MPD using the R packages 'ape'
125   and 'picante' – MPD is simply the mean of the distances of all the species in a community along
126   the edges of the evolutionary tree. To generate null models we simulated the assembly of
127   communities into each spectral cluster (n = 999) and calculated the MPD in each cluster (Paradis
128   et al., 2004; Kembel et al., 2010). We then compared the actual MPD for each cluster with the
129   simulated null distribution.